Hurricane Pattern Validation

The following project is an Exploratory Data Analysis of hurricanes data in the North Atlantic from years 1980 to 2019

Odaiclet Piccinini

9/29/2020

1) Data preparation:

The data contained in this report comes from the International Best Track Archive for Climate Stewardship (IB-TrACS) website (link here):

Specifically, the dataset is part of IBTrACS v04, the file is named ibtracs.NA.list.v04r00.csv (direct download here). This data has records of storms and hurricanes since 1980 which was considered the modern era of satellite observations. For the purpose of this report we will be focusing only on the Hurricanes in the North Atlantic.

Before going into the exploration analysis, it is worth noting the considerations that taken in this report before making assumptions about the data. See below:

Considerations:

- Recent advances in technology and research capabilities might be responsible for a possible increase in the amount of data available than there was during the 80's and 90's. The data comes from the originating agency representing units historically used by the community.
- As mentioned on this dataset's documentation the "IBTrACS ignores most names that include digits or abbreviations". Therefore, unnamed storms complicate the analysis of our data since we would not know exactly how to exactly classify. So, only *named* storms are considered to explore **all claims** in this report.

With such considerantions in mind we can continue our exploration.

```
# vector of names for first 16 columns to be used to name the columns of our Hurricanes data
col_names <- c("SID","SEASON", "NUMBER", "BASIN",
               "SUBBASIN", "NAME","ISO_TIME", "NATURE",
               "LAT", "LON", "WMO_WIND", "WMO_PRES", "WMO_AGENCY",
              "TRACK_TYPE", "DIST2LAND", "LANDFALL"
               )
```

```
# vector of data-types for first 16 columns to be used to assign the data types of the columns in our H
col_types <- c( "character", "integer", "integer", "character",
               "character", "character", "character", "character",
               "double", "double", "integer", "integer", "character",
               "character", "integer", "integer"
               )
```

```r
# Storing my data
# Data is the relevant data imported from the .csv file
dat <- read.csv(
  # the name of the file in my current directory
  file = "ibtracs.NA.list.v04r00.csv",
  #defying the data types of the columns
  colClasses = c(col_types, rep("NULL", 147)),
  # set to False to treat strings as string types
  stringsAsFactors = FALSE,
  # we're not interested in hurricanes before 1979
  skip = 86272,
  # Missing values are encoded as MM and should not
  # appear in the final product
  na.strings = " ",
                )
# renaming columns using vector col_names
colnames(dat) <- col_names

# adding a month column to table dat
dat$MONTH <- as.numeric(substr(dat$ISO_TIME, 6, 7))
# display structure of data
str(dat, vec.len = 1)
```

```
## 'data.frame':    36069 obs. of  17 variables:
##  $ SID       : chr  "1980199N31284" ...
##  $ SEASON    : int  1980 1980 ...
##  $ NUMBER    : int  49 49 ...
##  $ BASIN     : chr  "NA" ...
##  $ SUBBASIN  : chr  "NA" ...
##  $ NAME      : chr  "NOT_NAMED" ...
##  $ ISO_TIME  : chr  "1980-07-17 00:00:00" ...
##  $ NATURE    : chr  "TS" ...
##  $ LAT       : num  30.5 ...
##  $ LON       : num  -76.5 ...
##  $ WMO_WIND  : int  20 NA ...
##  $ WMO_PRES  : int  NA NA ...
##  $ WMO_AGENCY: chr  "hurdat_atl" ...
##  $ TRACK_TYPE: chr  "main" ...
##  $ DIST2LAND : int  390 382 ...
##  $ LANDFALL  : int  379 371 ...
##  $ MONTH     : num  7 7 ...
```

Research Claims

Using Exploratory Data Analysis techniques this report either confirms, denies, or clarifies each of the following claims.

- A) A typical hurricane season (during a calendar year) runs from June through November, but occasionally storms form outside those months.
- B) A typical year has 12 named storms, including six hurricanes of which three become major hurricanes (category 3, 4, and 5).

- C) September is the most active month (where most of the hurricanes occur), followed by August, and October.
- D) During the analyzed period (1980-2019), no hurricanes made U.S. landfall before June and after November.

To identify whether these claims are true or not, we must clarify that we are considering *Hurricanes* as Dr. Sanchez defines them on his textbook "*Computing With Data*". He says that "Only tropical cyclones that form over the Atlantic Ocean or eastern Pacific Ocean are called "hurricanes."" With his definition in mind, anytime that we refer to hurricanes, we have filtered our data by different categorical and quantitative data:

- The **SEASON** corresponds to the year in which the storm was recorded and it is a numeric type. We will filter by this variable on all of the claims.
- The **NAME** variable is the name of the storm given by respective agency and it has character type. In the entire report we will only use *named* storms or hurricanes.
- The **BASIN** variable is the position of all storms. In our dataset there are two unique regions present. One is the North Atlantic and the other one corresponds to Eastern North Pacific. In this report only the BASIN equal to "NA" are used.
- The **MONTH** is the number of the month in which the storm happened and its a numeric type. Throughout this report we will filter out by 1 or more months to explore in depth what happened during those months.
- The **WMO_WIND** is the maximum sustained speed assigned by the respective WMO agency. We will filter by this variable when we need to make a distinction between storms and hurricanes.
- The **LAT** and **LON** are the latitude of the mean position of the storm. The latitude can be a positive position above North of the Equator or negative position South of the Equator. The longitude can be the positive position East of the Meridian or negative position West of the Meridian. We will use the values in this variable for investigating claim D.
- The **LANDFALL** variable is the minimum distance of the storm to the land. When it equals 0 it represents the landfall within 6 hours. We will use the values in this variable for investigating claim D.

Having stated these parameters we must proceed with our exploration of this dataset.

Claim A)

"A typical hurricane season (during a calendar year) runs from June through November, but occasionally storms form outside those months."

Part 1)

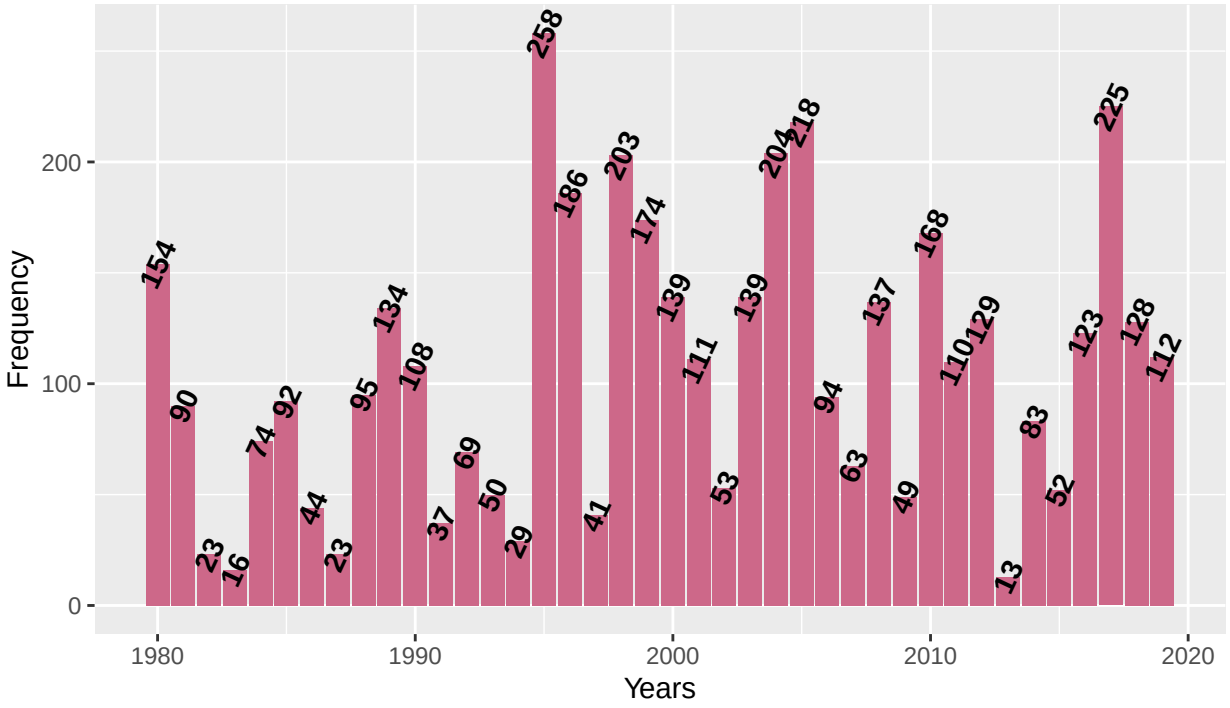
To explore the validity of this claim, data from relevant variables: SEASON, WMO_WIND, and BASIN are selected. Only the season from 1980 and 2019 are kept.

```
# Make a barplot for the frequency of the the storms for each entire year, per month.
# Bar plot to see how many storms happened since 1980 until 2019
storms_since_1980 <- dat %>%
   # select the variable season in the storms data
   select(SEASON, WMO_WIND, BASIN) %>%
   filter(SEASON  >= 1980 & SEASON <= 2019) %>%
```

```
   # filter by hurricanes
   filter(WMO_WIND >= 64) %>%
   # filter storms in the North Atlantic
   filter(BASIN == "NA") %>%
   # group all years
   group_by(SEASON) %>%
   # count how many of each year
      count(SEASON)
```

```
# ------------------# Details of the chart---------------------- #
ggplot(data = storms_since_1980, aes(x = SEASON, y = n)) +
  # Bar plot details
  geom_col(aes(x = SEASON), width = 0.9, fill = "palevioletred3") +
  # Frequency of storms
  geom_text(aes(label = n), angle = 65, fontface = "bold") +
  # set the range of the x axis to 1 - 12 corresponding Months of the year
  # set the Y and X labels
  xlab("Years") + ylab("Frequency") +
  # identify other important labels
  labs(
      # Give a title to the chart
      title = "            Fig 1.1 Storms from 1980 to 2019",
      # provide a descriptive footnote for the figure
      caption = "Bar Plot of Hurricanes Recordings from 1980 to 2019 taken from data reported on the In
      Best Track Archive for Climate Stewardship (IB- TrACS) website. The X axis corresponds to the Yea
      is the Number of detections of hurricanes in which these hurricanes happened.")
```

Fig 1.1 Storms from 1980 to 2019



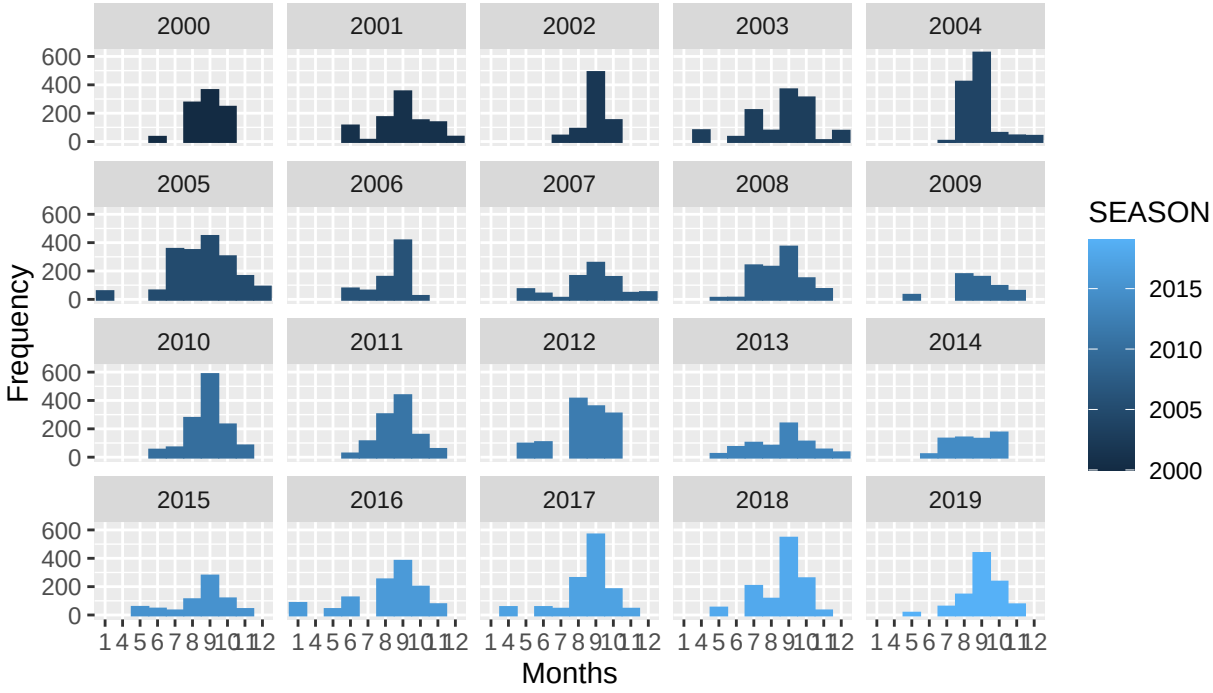
Bar Plot of Hurricanes Recordings from 1980 to 2019 taken from data reported on the International Best Track Archive for Climate Stewardship (IB− TrACS) website. The X axis corresponds to the Years and the Y axis is the Number of detections of hurricanes in which these hurricanes happened.

```
# Table with the storms from the last 20 years for faceted bar plot
hurricanes_last_20_years <- dat %>%
   # Filter storms from 2000 to 2019
   filter(SEASON >= 2000 & SEASON <= 2019)

# ------------------# Details of the chart---------------------- #
ggplot(data = hurricanes_last_20_years, aes(x = factor(MONTH), color = SEASON)) +
      # Make bar plots of the Years in the x axis and the count on the Y axis
      geom_bar(aes(fill = SEASON)) +
      # create a figure with the bar plots of the past 20 years
      facet_wrap(~ SEASON, shrink = FALSE) +
      # set the range of the x axis to 1 - 12 corresponding Months of the year
      # set the Y and X labels
      xlab("Months") + ylab("Frequency") +

    # Give a title to the chart
      ggtitle("          Fig 1.2 Hurricane Season Trend by Month
              Faceting by Year For The Last 20 Years.") +
      # provide a descriptive footnote for the figure
      labs(caption = "Each bar plot in the figure corresponds to one year of hurricanes. The X axis refl
      Months of the year (1 - 12) and the Y axis corresponds to the Frequency of Hurricanes")
```

Fig 1.2 Hurricane Season Trend by Month
Faceting by Year For The Last 20 Years.



Each bar plot in the figure corresponds to one year of hurricanes. The X axis reflects the Months of the year (1 − 12) and the Y axis corresponds to the Frequency of Hurricanes

My response to claim A:

After exploring the data in the Hurricanes table, I chose to focuse my attention on the period of the last 20 years (2000-2019). What lead me to this decision was that looking at Fig 1.1. I noticed that there were more Hurricanes relative to the years prior to 2000. The plot seems to show higher lows which results in an upward trend of hurricanes per year. The higher bars might also be attributed to the advances in technology of the past 25 years, which has allowed better, more reliable, and larger amounts of data to be collected. Therefore, this is not enough evidence to suggest that there will be more hurricanes in the future. Instead it could just be a feature to the information era in which the data was collected.

Keeping the above considerations in mind, we conclude in favor of this claim given the large volume of hurricane activity during the months of June to November as apparent in Fig 1.2. Such results suggest that there is a pattern on the appearance of Hurricanes during these months. So it is reasonable to conclude that the appearance of storms during these months is typical. In contrast, there seems to be little to no Hurricane activity during the months outside this range. This behavior is consistent with the part of the claim that expresses that ocassionaly storms form in the months outside of this range.

Claim B)

"A typical year has 12 named storms, including six hurricanes of which three become major hurricanes (category 3, 4, and 5)."

Part 2)

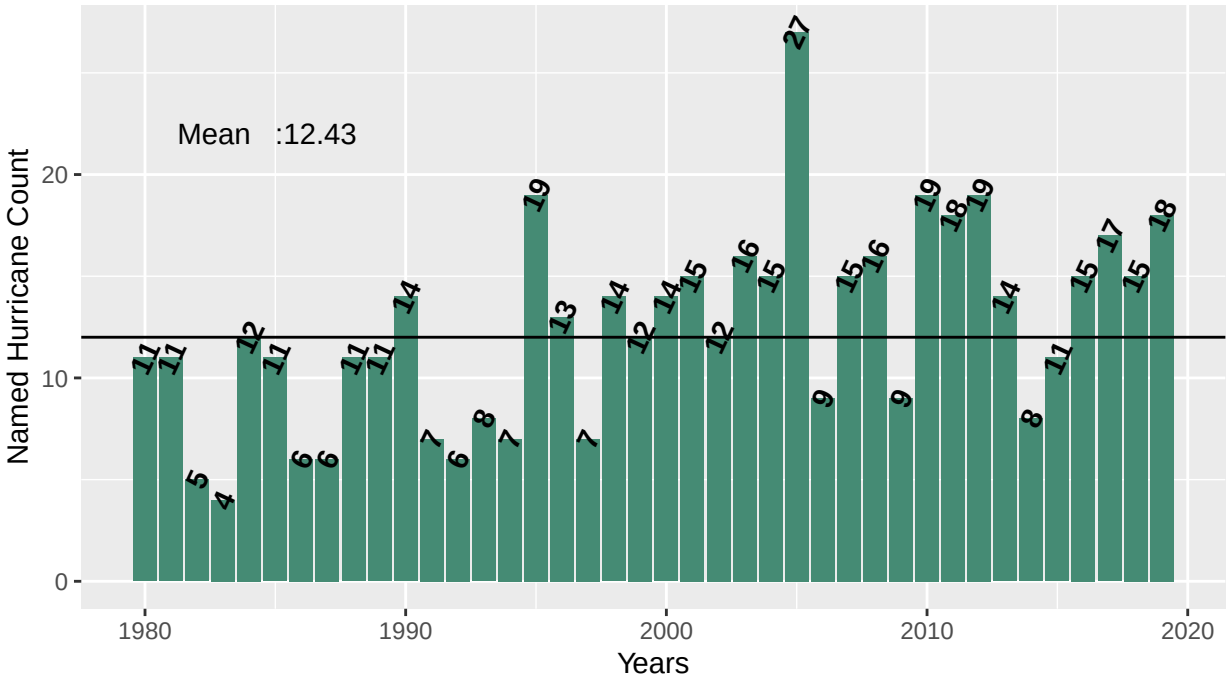
In order to explore this claim further, it is important to dive in the unique number of storms per year.

- **Below are:
- A plot with the number of storms per year.
- Calculations of the average number of storms per year.

```
# calculate the average number of storms per year
avg_storms_num <- typical_storms_per_year %>%
   # group by year
   group_by(SEASON) %>%
   # produce a summary table
   summary()
# extract the mean value
storms_median <- avg_storms_num[10]
```

```
# ------------------# Details of the chart---------------------- #
ggplot(data = typical_storms_per_year, aes(x = SEASON, y = n)) +
   # Bar plot details
   geom_col(aes(x = SEASON), width = 0.9, fill = "aquamarine4") +
   # Frequency of hurricanes +
   annotate(geom = "text", y = 22, x = 1985, label = storms_median) +
   geom_text(aes(label = n), angle = 65, fontface = "bold") +
   # set the range of the x axis to 1 - 12 corresponding Months of the year
   # set the Y and X labels
   xlab("Years") + ylab("Named Hurricane Count") +
   # produce a horizontal line with the number 12 (typical number of hurricanes named per year)
   geom_hline(aes(yintercept = 12)) +
   # identify other important labels
   labs(
      # Give a title to the chart
      title = "     Fig 2.1. Named Hurricanes from 1980 to 2019",
      # provide a descriptive footnote for the figure
      caption = "Bar Plot of Hurricanes from 1980 to 2019 taken from data reported on the International
      Archive for Climate Stewardship (IB- TrACS) website's. The X axis corresponds to the Years and th
      corresponds to the Number of Unique Hurricanes. The horizontal line is the value of 12 (believed
      average number of hurricanes per year)")
```

Fig 2.1. Named Hurricanes from 1980 to 2019



Bar Plot of Hurricanes from 1980 to 2019 taken from data reported on the International Best Track Archive for Climate Stewardship (IB− TrACS) website's. The X axis corresponds to the Years and the Y axis corresponds to the Number of Unique Hurricanes. The horizontal line is the value of 12 (believed to be the average number of hurricanes per year)

Important background information to consider:

To address the section of claim B that makes reference to the categories of Hurricanes, we are using the **Saffir-Simpson Hurricane Wind Scale** found on our class' textbook, *Computing With Data* by UC Berkeley Department of Statistics lecturer, Dr. Gaston Sanchez.

The **Saffir-Simpson Hurricane Wind Scale** ranks Hurricanes that sustained wind speed (in knots) as follows:

- category 1: 64-82 kt
- category 2: 83-95 kt
- category 3: 96-112 kt
- category 4: 113-136 kt
- category 5: 137 kt or higher

Additionally, according to the *National Hurricane Center and Central Pacific Hurricane Center* and the *National Oceanic And Atmosepheric Administration*, major hurricanes are those who,

> "Reaching Category 3 and higher are considered major hurricanes because of their potential for significant loss of life and damage"

For more information, please visit source documentation page here.

Having established these wind speed parameters, we can explore how many hurricanes became major hurricanes since 1980 until 2019. It is our hope that this data can illustrate how many of those were present in each year.

```r
typical_storms_max_wind <- dat %>%
   filter(NAME != "NOT_NAMED") %>%
   filter((SEASON >= 1980 & SEASON <= 2019)) %>%
   filter(BASIN == "NA") %>%
   filter(WMO_WIND != "NA") %>%
   group_by(NAME, SEASON) %>%
   summarise(max_wind = max(WMO_WIND)) %>%
   arrange(desc(max_wind))
```

```
## `summarise()` regrouping output by 'NAME' (override with `.groups` argument)
```

Identifying Major Hurricanes for further observations:

To identify the number of major hurricanes, averages of wind speed are not considered. The reason to this is because not all storms reach category 3, 4, or 5 so the averages would not accurately reflect those that did reach such category. To account for that, we are using the maximum values of wind speed sustained by a hurricane instead.

- **Category 3 Hurricanes**

```r
# Major Hurricane categories breakdown
# Category 3 storms sustain a wind speed of 111-129 mph; 96-112 kt; 178-208 km/h
category_3_storms <- typical_storms_max_wind %>%
   filter(max_wind  >= 96 & max_wind <= 112) %>%
   group_by(SEASON) %>%
   distinct(NAME)
```

- **Category 4 Hurricanes**

```r
# Category 4 storms sustain a wind speed of 130-156 mph; 113-136 kt; 209-251 km/h
category_4_storms <- typical_storms_max_wind %>%
   filter(max_wind  >= 113 & max_wind  <= 136) %>%
   group_by(SEASON) %>%
   distinct(NAME)
```

- **Category 4 Hurricanes**

```r
# Category 5 storms sustain a wind speed of 157 mph or higher; 137 kt or higher; 252 km/h or higher
category_5_storms <- typical_storms_max_wind %>%
   filter(max_wind  >= 137) %>%
   group_by(SEASON) %>%
   distinct(NAME)
```

In the next section, we are investigating the the number of major hurricanes that happened per year. Therefore, we will be collecting data to find the averages. Studying the averages of each category by year allows us to get an overall picture of what our data represents.

```r
hurricanes_count <- dat %>%
   filter(NAME != "NOT_NAMED") %>%
   filter((SEASON >= 1980 & SEASON <= 2019)) %>%
   filter(WMO_WIND != "NA") %>%
   filter(WMO_WIND >= 64) %>%
   group_by(SEASON) %>%
   distinct(NAME) %>%
   count()

# Count all unique major hurricanes per category
# Category 3 Hurricanes count
category_3_count <- count(category_3_storms)
# Category 4 Hurricanes count
category_4_count <- count(category_4_storms)
# Category 5 Hurricanes count
category_5_count <- count(category_5_storms)
```

```r
# create a data frame with the sequence of years from 1980 to 2019
years <- data.frame(
   # use the seq function to create sequence of years by increments of 1 year
   SEASON = seq(1980, 2019, by = 1)
)

# join the count of hurricane category 3 by the year
hurricanes_cat_3_n <- full_join(years, category_3_count,
                                by = "SEASON",
                                copy = TRUE)
   # rename variable n with more relevant variable CAT 3
hurricanes_cat_3 <- rename(hurricanes_cat_3_n, CAT_3 = n)

# join the count of hurricane category 4 by the year with the table containing cat 3 hurricanes
hurricanes_cat_4_n <- full_join(hurricanes_cat_3, category_4_count,
                                by = "SEASON", copy = TRUE)
   # rename variable n with more relevant variable CAT 4
hurricanes_cat_4 <- rename(hurricanes_cat_4_n, CAT_4 = n)

# join the count of hurricane category 5 by variable year with the table containing cat 3 and 4 hurrica
hurricanes_cat_5_n <- full_join(hurricanes_cat_4, category_5_count,
                                by = "SEASON", copy = TRUE)
   # rename variable n with more relevant variable CAT 5
hurricanes_all_cats <- rename(hurricanes_cat_5_n, CAT_5 = n)

# Data frame of all major hurricane per year per category totaled by year
num_major_h_per_year <- hurricanes_all_cats %>%
   # replace missing values of CAT 3, CAT 4, CAT 5 with the 0 value for sum
   replace(is.na(.), 0) %>%
   # Indicate that next column will do rowise operations
   rowwise() %>%
   # Add a new column named total with the sum of all major hurricanes per year
   mutate(major_hurricanes = sum(c_across(CAT_3:CAT_5)))

# New data frame of summaries, including table named total_hurricanes, containing the count of all stor
summary_of_hurricanes <- full_join(num_major_h_per_year, hurricanes_count,
```

```
                                    by = "SEASON", copy = TRUE)
   # naming column n with new name "total_hurricanes"
summary_of_hurricanes <- summary_of_hurricanes %>%
   rename(total_storms = n)
```

Fig 2.3 Summary table of Hurricanes and Major Hurricanes by Year

```
summary_of_hurricanes
```

```
## # A tibble: 40 x 6
## # Rowwise:
##    SEASON CAT_3 CAT_4 CAT_5 major_hurricanes total_storms
##     <dbl> <dbl> <dbl> <dbl>            <dbl>        <int>
##  1   1980     1     0     1                2            9
##  2   1981     2     1     0                3            7
##  3   1982     0     1     0                1            2
##  4   1983     1     0     0                1            3
##  5   1984     0     1     0                1            5
##  6   1985     2     1     0                3            7
##  7   1986     0     0     0                0            4
##  8   1987     1     0     0                1            3
##  9   1988     0     2     1                3            6
## 10   1989     0     1     1                2            7
## # ... with 30 more rows
```

```
# displaying summary table
detailed_summary <- summary(summary_of_hurricanes)

# Extracting "mean" values from table of summaries
mean_CAT3_hurricanes = detailed_summary[10]
mean_CAT4_hurricanes = detailed_summary[16]
mean_CAT5_hurricanes = detailed_summary[22]
mean_major_hurricanes = detailed_summary[28]
mean_storms = detailed_summary[34]

sum_cat3_h <- sum(summary_of_hurricanes$CAT_3)
sum_cat4_h <- sum(summary_of_hurricanes$CAT_4)
sum_cat5_h <- sum(summary_of_hurricanes$CAT_5)
sum_majo_h <- sum(summary_of_hurricanes$major_hurricanes)
sum_all_h <- sum(summary_of_hurricanes$total_storms)

# data frame with means
means_table <- data.frame(
   type = c("Category 3 Hurricanes", "Category 4 Hurricanes", "Category 5 Hurricanes", "Major Hurricanes
   average_occurrance_typical_year = c(mean_CAT3_hurricanes, mean_CAT4_hurricanes, mean_CAT5_hurricanes
   total = c(sum_cat3_h, sum_cat4_h, sum_cat5_h, sum_majo_h, sum_all_h)
)
# Print the means table below
```

Fig 2.4: Table of the Average Number of Typical Hurricane Occurrances in a Year

```
means_table
```

```
##                    type average_occurrance_typical_year total
## 1 Category 3 Hurricanes                  Mean   :1.075      43
## 2 Category 4 Hurricanes                   Mean   :1.15      46
## 3 Category 5 Hurricanes                  Mean   :0.475      19
## 4      Major Hurricanes                    Mean   :2.7     108
## 5             All Storms                  Mean   : 6.65     266
```

My response to claim B:

"A typical year has 12 named storms, including six hurricanes of which three become major hurricanes (category 3, 4, and 5)."

> As previously stated, storms become hurricanes after they have sustained a wind speed of 64 knots or higher. Considering this fact, we must observe on Fig. 2.1 that in the entirety of the period of 1980 to 2019, the average number of *named* storms that occurred in one year was *12.43*. From this bar plot, we can see that the number of storms below average is greater than the number of storms above average.
>
> To illustrate if, from such storms, "six were hurricanes"; we must look at Fig. 2.4. This table provides us with a summary of the average occurrance of Hurricanes. According to our Fig.2.4, the average number of Hurricanes that occur in a given year is 6.65. From these Hurricanes, the average number of huricanes that become category 3, 4, and 5 (major hurricanes) is 2.7. Considering these averages, Claim B is reasonable in the sense that 12 storms are typical of a hurricane season, but seven storms are more realistic to be hurricanes instead of six. Also, it is reasonable that from the typical number of hurricanes 3 of them are major hurricanes. This can be supported by the values that we have obtained because both averages are close to or slightly above to the number stated in claim B. For this reason, we are in favor of this claim.

Claim C)

"September is the most active month (where most of the hurricanes occur), followed by August, and October."

Part 3)

As we have previously mentioned, for a storm to be considered a hurricane it must have sustained winds that reached at least 64 knots. Therefore and to investigate this claim further it seems a reasonable approach to produce bar plots of hurricane data recorded faceted by months (only listing August, September, and October) for the last 20 years.

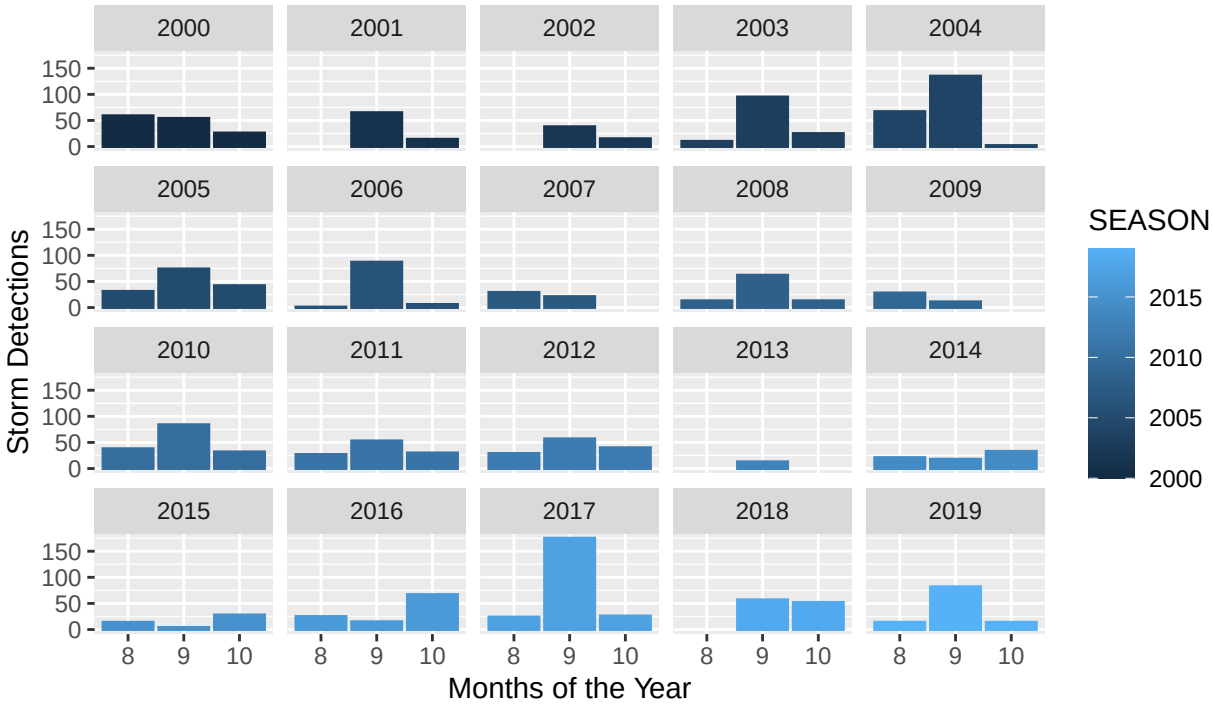
```
most_active_month <- dat %>%
   # Exclude the storms that are not named
   filter(NAME != "NOT_NAMED") %>%
   # Extract only hurricanes in the past 20 years
   filter((SEASON >= 2000 & SEASON <= 2019)) %>%
   # filter out hurricanes with no wind speed values and select only those
```

```
   # with wind speed values above 64 kt.
   filter(WMO_WIND != "NA" & WMO_WIND >= 64) %>%
   # filter only by North Atlantic hurricanes
   filter(BASIN == "NA") %>%
   # filter only the months of August, September, and October
   filter(MONTH >= 8 & MONTH <= 10) %>%
   # Select only relevant columns of data
   select(SEASON, NAME, WMO_WIND, MONTH) %>%
   # Group data by core variables SEASON and MONTH
   group_by(SEASON, MONTH)
```

```
# ------------------# Details of the chart---------------------- #
# using ggplot to form a bar plot of the data collected of the most active months
# and facet by year, while factoring by month
ggplot(data = most_active_month, aes(x = factor(MONTH), color = SEASON)) +
      # Make barplots of the Years in the x axis and the count on the Y axis
      geom_bar(aes(fill = SEASON)) +
      # create a figure with the bar plots of the past 20 years
      facet_wrap(~ SEASON, shrink = FALSE) +
      # Set the Y and X labels
      # The x axis are months 8 through 10, corresponding to August, September, and October.
      xlab("Months of the Year") + ylab("Storm Detections") +

    # Give a title to the chart
      ggtitle("Fig. 3.1 Most Active Months in a Typical Hurricane Season (past 20 Years)") +
      # provide a descriptive footnote for the figure
      labs(caption = "Each bar plot in the figure corresponds to one year of hurricanes. The X axis refl
      August (8), September (9), and October (10). The Y axis corresponds to the Frecuency (For this pl
      we take frequency as the total number of Hurricanes detections that month.")
```

Fig. 3.1 Most Active Months in a Typical Hurricane Season (past 20 Years)



Each bar plot in the figure corresponds to one year of hurricanes. The X axis reflects the Months of ıgust (8), September (9), and October (10). The Y axis corresponds to the Frecuency (For this plot, we take frequency as the total number of Hurricanes detections that month.

- **August data for all years:**

```
august_data <- dat %>%
   filter(NAME != "NOT_NAMED") %>%
   # Extract only hurricanes in the past 20 years
   filter((SEASON >= 1980 & SEASON <= 2019)) %>%
   # filter out hurricanes with no wind speed values and select only those
   # with wind speed values above 64 kt.
   filter(WMO_WIND != "NA" & WMO_WIND >= 64) %>%
   # filter only by North Atlantic hurricanes
   filter(BASIN == "NA") %>%
   # filter only the months of August, September, and October
   filter(MONTH == 8) %>%
   # Select only relevant columns of data
   select(SEASON, NAME, WMO_WIND, MONTH) %>%
   # Group data by core variables SEASON and MONTH
   group_by(SEASON, MONTH) %>%
   count()
# calculate the average number of August hurricane detections for all years
mean_august <- mean(august_data$n)
# calculate the total number of August hurricane detections for all years
sum_august <- sum(august_data$n)
```

- **September data for all years:**

```r
september_data <- dat %>%
   filter(NAME != "NOT_NAMED") %>%
   # Extract only hurricanes in the past 20 years
   filter((SEASON >= 1980 & SEASON <= 2019)) %>%
   # filter out hurricanes with no wind speed values and select only those
   # with wind speed values above 64 kt.
   filter(WMO_WIND != "NA" & WMO_WIND >= 64) %>%
   # filter only by North Atlantic hurricanes
   filter(BASIN == "NA") %>%
   # filter only the months of August, September, and October
   filter(MONTH == 9) %>%
   # Select only relevant columns of data
   select(SEASON, NAME, WMO_WIND, MONTH) %>%
   # Group data by core variables SEASON and MONTH
   group_by(SEASON, MONTH) %>%
   count()
# calculate the average number of September hurricane detections for all years
mean_september <- mean(september_data$n)
# calculate the total number of September hurricane detections for all years
sum_september <- sum(september_data$n)
```

- **October data for all years:**

```r
# find what happens in October for all years
october_data <- dat %>%
   filter(NAME != "NOT_NAMED") %>%
   # Extract only hurricanes in the past 20 years
   filter((SEASON >= 1980 & SEASON <= 2019)) %>%
   # filter out hurricanes with no wind speed values and select only those
   # with wind speed values above 64 kt.
   filter(WMO_WIND != "NA" & WMO_WIND >= 64) %>%
   # filter only by North Atlantic hurricanes
   filter(BASIN == "NA") %>%
   # filter only the months of August, September, and October
   filter(MONTH == 10) %>%
   # Select only relevant columns of data
   select(SEASON, NAME, WMO_WIND, MONTH) %>%
   # Group data by core variables SEASON and MONTH
   group_by(SEASON, MONTH) %>%
   count()
# calculate the average number of September hurricane detections for all years
mean_october <- mean(october_data$n)
# calculate the total number of September hurricane detections for all years
sum_october <- sum(october_data$n)
```

```r
# Data frame with a summary of the average hurricane detections
# and the total number of hurricanes for all years since 1980.
most_active <- data.frame(
   month = c("August", "September", "October"),
   average_hurricanes = c(mean_august, mean_september, mean_october),
   sum_of_hurricanes = c(sum_august, sum_september, sum_october)
)
```

Fig. 3.2 Table of Average Number of Active Hurricanes

```
most_active
```

```
##       month average_hurricanes sum_of_hurricanes
## 1    August           28.96875               927
## 2 September           54.58974              2129
## 3   October           23.93548               742
```

My response to claim C:

"September is the most active month (where most of the hurricanes occur), followed by August, and October."

> As we can see on Table 3.1, during the past 20 years of recorded hurricanes, September showed to have a larger bar in 13 of the 20 years. This ilustrates that 65% of the time, September was more active than the remaining 2 months. But this is not enough evidence to validate or negate this claim.
>
> Therefore, it seemed necesary and beneficial to calculate the averages of the number of hurricane activity per year by month. That way we can make stronger associations between our variables. On table 3.2, we have listed a summary of the average hurricane activity and the total number of hurricane recordigns per month for all years. In both columns we can see that September's values are more than double compared to the other values at approximately 54.59. August and October averaged approximately 28.97 and 23.94 respectively. Therefore, we are in favor of this claim.

Claim D)

"During the analyzed period (1980-2019), no hurricanes made U.S. landfall before June and after November."

To begin exploring this claim it is necessary to remember what we mean by LANDFALL, LAT and LON.

- The **LANDFALL** as defined by the iBTracs is the "minimun distance to land over next 3 hours (= 0 means landfall)".
- The **LAT** and **LON** corresponds to the latitude position and logitude position respectively. To put it in context, the latitude and longitude of the United States are 37.0902° North, 95.7129° West. You might have noticed that these are the coordinate positions of the country. Typically these coordinates are expressed in degrees. Since the United states is ~95° West. This number will be **negative** in our data set. The reason is because it refers to the West side of the Meridian.

```
# Create a table of the hurricanes that made landfall in the US
# during 1980 through 2019
hurricanes_and_landfall <- dat %>%
   # Extract only hurricanes in the past 20 years
   filter((SEASON >= 1980 & SEASON <= 2019)) %>%
   # # Extract US hurricanes by coordinates of Latitude and longitude
   filter(LAT >= 25 & LAT <= 50) %>%
```

```
  filter(LAT > -65) %>%
  # filter out the unnamed hurricanes
  filter(NAME != "NOT_NAMED") %>%
  # filter by core variable WMO_WIND: filter out hurricanes with
  #no wind speed values and select only with wind speed values above 64 kt.
  filter(WMO_WIND != "NA" & WMO_WIND >= 64) %>%
  # filter core variable BASIN by North Atlantic hurricanes
  filter(BASIN == "NA") %>%
  # Select only the relevant columns of data
  select(SEASON, NAME, WMO_WIND, MONTH, LANDFALL) %>%
  # Filter by core variable LANDFALL to find which hurricanes made US landfall
  filter(LANDFALL == 0) %>%
  # group by core variable season
  group_by(SEASON) %>%
  # count the grouped years
  count(SEASON)
```

Fig. 4.1. Table of years when Hurricanes made landfall in the US

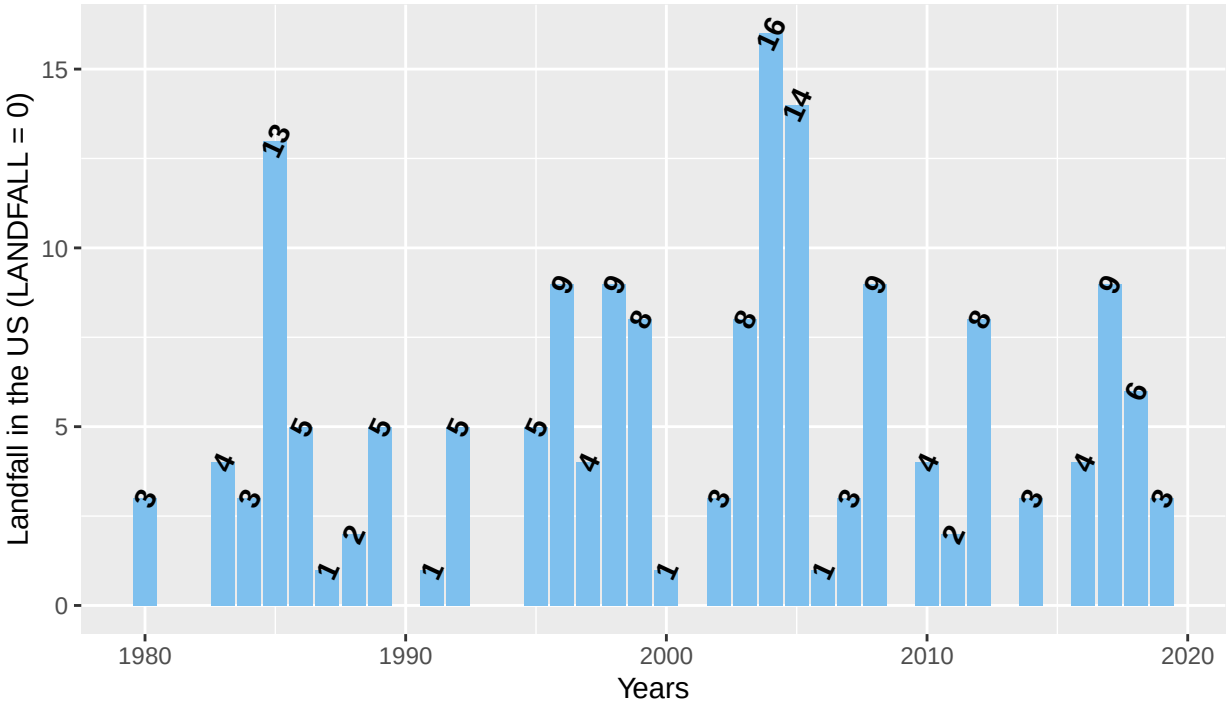
```
hurricanes_and_landfall
```

```
## # A tibble: 31 x 2
## # Groups:   SEASON [31]
##    SEASON     n
##     <int> <int>
##  1   1980     3
##  2   1983     4
##  3   1984     3
##  4   1985    13
##  5   1986     5
##  6   1987     1
##  7   1988     2
##  8   1989     5
##  9   1991     1
## 10   1992     5
## # ... with 21 more rows
```

```
# ------------------# Details of the chart---------------------- #
ggplot(data = hurricanes_and_landfall, aes(x = SEASON, y = n)) +
  # Bar plot details
  geom_col(aes(x = SEASON), width = 0.9, fill = "skyblue2") +
  # Frequency of storms
  geom_text(aes(label = n), angle = 65, fontface = "bold") +
  # set the range of the x axis to 1 - 12 corresponding Months of the year
  # set the Y and X labels
  xlab("Years") + ylab("Landfall in the US (LANDFALL = 0)") +
  # identify other important labels
  labs(
      # Give a title to the chart
      title = "Fig. 4.2. Frequency of Hurricane Landfall in the United States",
      # provide a descriptive footnote for the figure
```

```
      caption = "Bar Plot of Hurricanes that made landfall in the United States, from 1980 to 2019
      The X axis corresponds to the Years since 1980 and the Y axis corresponds to
      the Frequency of hurricanes that made US landfall that year.")
```

Fig. 4.2. Frequency of Hurricane Landfall in the United States

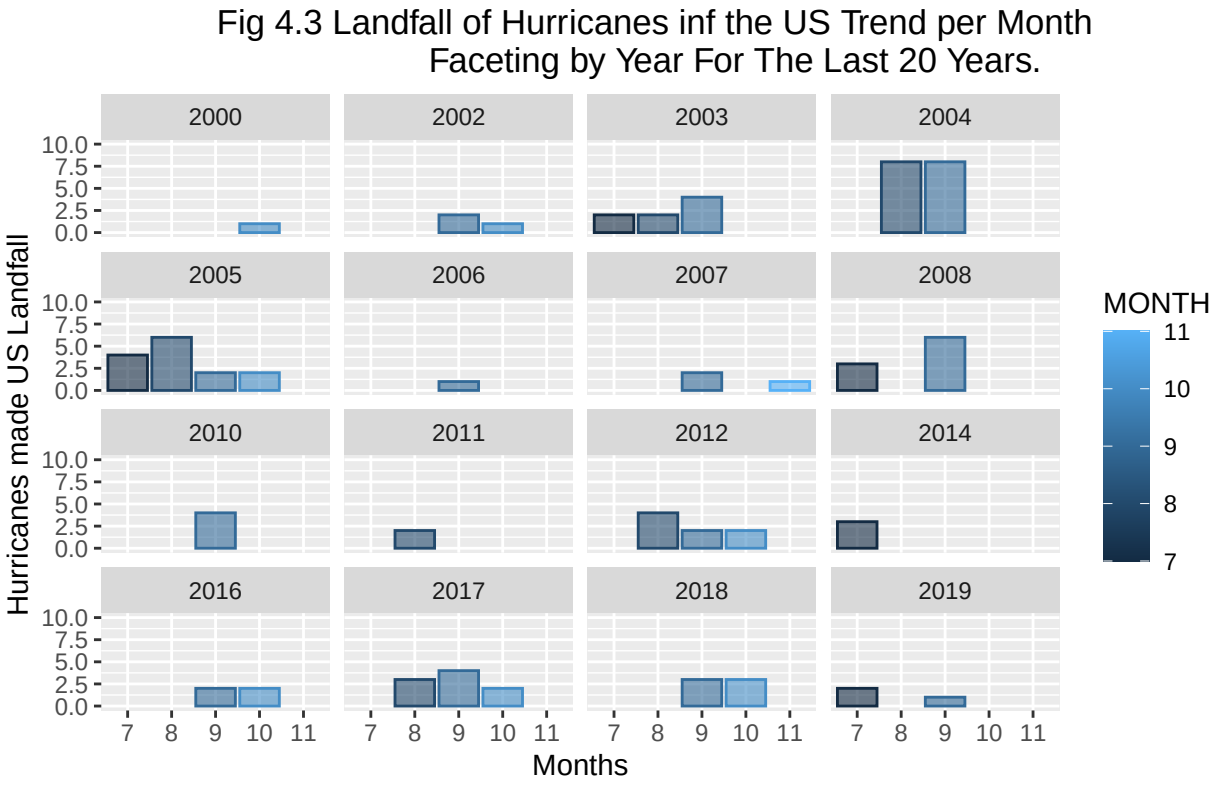


Bar Plot of Hurricanes that made landfall in the United States, from 1980 to 2019
The X axis corresponds to the Years since 1980 and the Y axis corresponds to
the Frequency of hurricanes that made US landfall that year.

```
# ------------------# Details of the chart---------------------- #
# Table with the storms from the last 20 years
landfalls_last_20_years <- dat %>%
   # Extract hurricanes by core variable SEASON for the past 20 years
   filter((SEASON >= 2000 & SEASON <= 2019)) %>%
   # filter core variable BASIN by North Atlantic hurricanes
   filter(BASIN == "NA") %>%
   # Extract US hurricanes by coordinates of Latitude and longitude
   filter(LAT >= 25 & LAT <= 50) %>%
   filter(LAT > -65) %>%
   # filter out the unnamed hurricanes
   filter(NAME != "NOT_NAMED") %>%
   # filter by core variable WMO_WIND: filter out hurricanes with
   # no wind speed values and select only with wind speed values above 64 kt.
   filter(WMO_WIND != "NA" & WMO_WIND >= 64) %>%
   # hurricanes that made US landfall
   filter(LANDFALL == 0)
```

Fig. 4.3. Trend of Hurricane landfall in the US per Month, Faceting by Year For The Last 20 Years.

```
# ------------------# Details of the chart---------------------- #
ggplot(data = landfalls_last_20_years, aes(x = factor(MONTH), color = MONTH)) +
      # Bar plots of the Years in the x axis and the count on the Y axis
      geom_bar(aes(fill = MONTH),
               alpha = 0.6) +
      # create a figure with the bar plots of the past 20 years
      facet_wrap(~ SEASON, shrink = FALSE) +
      # set the range of the x axis to 1 - 12 corresponding Months of the year
      # set the Y and X labels
      xlab("Months") + ylab("Hurricanes made US Landfall") +
      ylim(0,10) +
    # Give a title to the chart
      ggtitle("            Fig 4.3 Landfall of Hurricanes inf the US Trend per Month
                                  Faceting by Year For The Last 20 Years.") +
      # provide a descriptive footnote for the figure
      labs(caption = "Each bar plot in the figure corresponds to one year of hurricanes. The X axis refl
      Months of the year and the Y axis corresponds to the number of Hurricanes that made landfall.")
```



Each bar plot in the figure corresponds to one year of hurricanes. The X axis reflects the Months of the year and the Y axis corresponds to the number of Hurricanes that made landfall.

```
# Create a table to find out exactly the earliest month and the latest month of hurricane landfall in t
since_1980_landfall <- dat %>%
   # Extract hurricanes by core variable SEASON for the past 20 years
   filter((SEASON >= 1980 & SEASON <= 2019)) %>%
```

```
   # filter core variable BASIN by North Atlantic hurricanes
   filter(BASIN == "NA") %>%
   # Extract US hurricanes by coordinates of Latitude and longitude
   filter(LAT >= 25 & LAT <= 50) %>%
   filter(LAT > -65) %>%
   # filter out the unnamed hurricanes
   filter(NAME != "NOT_NAMED") %>%
   # filter by core variable WMO_WIND: filter out hurricanes with
   # no wind speed values and select only with wind speed values above 64 kt.
   filter(WMO_WIND != "NA" & WMO_WIND >= 64) %>%
   # hurricanes that made US landfall
   filter(LANDFALL == 0)
```

```
# Data frame of the min and max months of hurricane landfall in the US
landfall_table <- data.frame(
   at = c("minimum", "maximum"),
   month = c(min(since_1980_landfall$MONTH), max(since_1980_landfall$MONTH))
)
```

- **Earliest and latest month when there was hurricane landfall in the US**

```
landfall_table
```

```
##        at month
## 1 minimum     6
## 2 maximum    11
```

My response to claim D:

To see if there where any hurricane landfalls in the US during 1980 and 2019, first we created a table with the count of hurricane landfalls per year. Then, we produced a bar plot with all data from the table in Fig 4.1. The visualization in Fig. 4.2, assured us that there where years in which there was landfall in the US and when there was not. After, we thought it was important to create a plot faceted by year, highligthing bar plots with the frequency in which hurricanes made landfall by month for the past 20 years. This data provided us with information trend of each month per year, which we were hoping privided enough evidence to give an opinion about the claim. However, it can be seen on out Fig. 4.3 that there are missing months in our plot. If we were only to present this plot, one could argue that it is no use because there might be data in the other months that we are not showing.

Therefore, it was decided to make a table with the minimun and maximun months in which hurricane landfall occurred. As it can be appreciated on Fig. 4.3, the earliest month in which landfall occurred was June (6), and the latest month was November (10). This alone seems to be enough evidence to support the claim stating that there were no landfalls outside of those months, more specifically, that there were no landfalls during January, February, March, April, May, or December. Therefore, we are in favor of this claim

Conclusions

The exploratory data analysis done allowed a deeper investigation of each claim, which illustrated a better picture for exploration. Claim A, is true because in all years the frequency of hurricanes

during June-November was larger. In addition, in claim B) the typical number of named storms was consistent with the average number of named storms for all years. Even though there was one more hurricane on average than what was stated (stated: 6, average: ~7), the same number of hurricanes became major hurricanes (stated: 3, Average: ~3). Additionaly, for claim C we are in favor of the claim, since september showed double the average of hurricanes per year than the other months(August and October). Finally on claim D, we decided to be in favor of the claim because our exploration sowed there were no hurricanes that made landfall in the US outside of the hurricane season (June - November). Besides a small rounding discrepancy on claim B, we are in favor of all of the four claims investigated in this report.

Video Submission:

A video that supports claim B is uploaded to my respective UC Berkeley Google Drive account and the link is here. Only UC Berkeley members can see this video.

In case that the link does not work, it is also uploaded on YouTube and the link is here